

# The International Journal of Digital Curation

## Issue 1, Volume 6 | 2011

### Towards Support for Long-Term Digital Preservation in Product Lifecycle Management

Wolfgang Wilkes, Jörg Brunsmann, Dominic Heutelbeck, Andreas Hundsdörfer,  
Matthias Hemmje, and Hans-Ulrich Heidbrink,  
University of Hagen,  
InConTec GmbH<sup>1</sup>

#### Abstract

Important legal and economic motivations exist for the design and engineering industry to address and integrate digital long-term preservation into product lifecycle management (PLM). Investigation revealed that it is not sufficient to archive only the product design data that are created in early PLM phases, but that preservation is needed for data that are produced during the entire product lifecycle including early and late phases. Data that are relevant for preservation consist of requirements analysis documents, design rationale, data that reflect experiences during product operation and also metadata like social collaboration context. In addition, the engineering environment itself, that contains specific versions of all tools and services, is a candidate for preservation. This paper takes a closer look at engineering preservation use case scenarios as well as PLM characteristics and workflows that are relevant for long-term preservation. Resulting requirements for a long-term preservation system lead to an Open Archival Information System (OAIS) -based system architecture and a proposed preservation service interface that respects the needs of the engineering industry.<sup>2</sup>

---

<sup>1</sup> InConTec GmbH: [http://www.incontec.de/ICT\\_Webpage\\_en.htm](http://www.incontec.de/ICT_Webpage_en.htm).

<sup>2</sup> This paper is based on the paper given by the authors at iPRES 2009; received January 2010, published March 2011.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.



## Introduction

The Sustaining Heritage Access through Multivalent ArchiviNg (SHAMAN) digital preservation project<sup>3</sup> investigates the needs of different application domains, which are known as Integration & Demonstration Subprojects (ISPs). The ISPs include:

- Memory institutions: scientific publishing in libraries and documents in governmental archives;
- Data resources used in e-Science applications;
- The industrial design and engineering industry.

The industrial design and engineering industry is characterised by a large number of heterogeneous design and development tools that are organised by product lifecycle management (PLM) and product data management (PDM) systems (SHAMAN, 2008). Those PLM and PDM systems provide services such as company-adapted workflows, data integration, and version and configuration management. A large number of heterogeneous data are created during execution of all PLM phases. Companies in the design and engineering industry consider long-term preservation of such product data as an additional asset to their PLM/PDM systems. As discussed by Heutelbeck, Brunsmann, Wilkes, and Hundsdörfer (2009), there are several important motivations for design and engineering companies to engage in digital long-term preservation. These include legal requirements defined by law, and contracts, as well as economic reasons, such as efficient reuse. For long-term preservation, the Open Archival Information System (OAIS; Consultative Committee for Space Data Systems, 2002) is widely accepted as a reference architecture. By analysing the design and engineering industry, this paper derives a number of requirements which have to be satisfied by digital preservation systems in order to be adoptable within state-of-the-art design and engineering environments.

This paper is organised as follows: we first give a short introduction to the OAIS reference architecture. We then lay out several use case scenarios for preservation system access in the engineering industry. We proceed by describing characteristics of PLM-based workflows and existing shortcomings of these implementations. From the described use case scenarios and shortcomings we derive requirements for long-term preservation systems in the engineering industry. Finally, we propose a high-level architecture and service interface for a long-term preservation system capable of supporting the needs of PLM systems.

## The Open Archival Information System Model

An OAIS is an archive consisting of an organisation of people and systems that have accepted the responsibility to preserve information and make it available for a Designated Community. The OAIS provides a conceptual reference architecture and does not specify a concrete implementation.

<sup>3</sup> SHAMAN Homepage: <http://shaman-ip.eu/shaman/>.



The **OAIS environment model** consists of the OAIS archive and three external actors.

- A *producer* is a person, organisation, or system that provides the digital information to be preserved.
- A *consumer* is a person, organisation, or system that accesses the OAIS to find and acquire preserved information. The Designated Community, which may be composed of multiple user communities, is an identified group of potential consumers who should be able to understand a particular set of information.
- The *management* are those persons or organisations who set the overall OAIS policies in the sense of management's responsibilities, and not the daily operational archive administration.

While investigating use cases and requirements, it is necessary to focus on the needs of the Designated Community, which will be done in the next section.

The **OAIS functional model** consists of six major functional entities. The Ingest functional entity accepts one or more Submission Information Packages (SIPs) and creates Archival Information Packages (AIPs), which it provides to Archival Storage for preservation. Ingest also sends Descriptive Information (DI) to the Data Management functional entity. Consumers interact with the Access function, which uses the DI to find the content information of interest. Access retrieves AIPs from Archival Storage and sends Dissemination Information Packages (DIPs) to the consumer. Administration oversees the day-to-day operation of the archive, and it receives advice from Preservation Planning on evolving strategies and mechanisms for preservation. All six functional entities are further broken down in the OAIS reference model.

OAIS-based preservation systems might use some of the following preservation methods:

- *Migration* is the continuous translation of data between data formats and systems;
- *Transformation* is the conversion of data from one format to another format that is assumed to be more preservable;
- *Emulation* (virtualisation) is the duplication of functionality of a system running on another system.

## Designated Community Use Case Scenarios in the Engineering Industry

Since the requirements for a preservation system are driven by the needs of the preservation system consumers, we study these use cases in more detail. The structures of use case scenarios for preserving engineering data follow the basic motivations of corporate preservation, as introduced by Heutelbeck et al. (2009). Here, we extend the motivations with concrete examples.

### *Legal Use Cases*

In legal scenarios, preservation is mandatory by either contract or law. Legal defense can be prepared using preserved-engineering data. Some examples are:

- A malfunction in a 40-year-old aircraft results in a crash which leads to serious injuries, or even death, of pilots and passengers. Investigating agencies demand access to preserved data from the manufacturing company.
- Malfunction in medical equipment can cause serious health problems for patients or can even lead to deaths.
- A malfunction in automotive electronics destroys other parts of a car and can result in an accident.

### *Economic Use Cases*

Economic use cases seek to increase the return on investment by lowering the cost for activities such as reuse, maintenance, refitting, and training. Examples are:

- As a consequence of a malfunction in a 40-year-old aircraft, unexpected downtimes of aircraft, spacecraft, or satellites occur that result in financial damage for the company. An engineer of the operating company will use the preserved product data to analyse a potential error.
- Products with a long lifetime are occasionally enhanced during their lifetime, either to introduce new functionality, or to adopt new requirements coming from customers or the market. Depending on the impact of these modifications on different areas of a product, it might be necessary to check either portions of the data or the total product data.
- A future engineer takes some parts (e.g. the enclosure) of an existing product and reuses it for a new product. Sophisticated search functionality supports him in finding the product data that fit his needs. Following this kind of approach the advantage is a shorter time to market for a new product and the reuse of parts which have already passed quality assurance tests, etc.
- A third-party provider for electronic equipment wants to develop and market add-on equipment for a company's product. He can selectively access the company product design data to investigate specification details and design recommendations for interfaces.
- A product recall happens as a result of malfunction. The engineering department has to check the product design data, fix the problems, and create a new revision of the product to pass it again to manufacturing.

### *Archive Consumers*

Based on these use case scenarios we can identify the actors (consumers, Designated Community) for accessing engineering data that are preserved on a long-term basis:

- Future engineers
- Investigators
- Other companies
- Governmental authorities
- Regulatory agencies

We can identify two major kinds of requirements from an access point-of-view:

- Access of the (native) engineering design data for reuse. The engineer needs full access to those parts of the design data which he wants to reuse. In electronic engineering, they are the logical design, simulation models, placement data for the physical design, specifications, etc;
- Access to representations of the design data, for example, PDF documents, and images of 2D drawings, which have been created from the native design data. These may be used to support service personnel, regulatory needs, or, for example, ISO9000<sup>4</sup> requirements.

## Product Lifecycle Management Characteristics

This section lays out the fundamental properties that exist in PLM-based environments. We describe characteristics that are inherent to the engineering industry and that affect requirements and in consequence the system architecture.

### *Overview*

Products in the aerospace, automotive, chemical and petroleum, electronics, energy and utilities, as well as shipbuilding industries, pass a product lifecycle that spans from phases such as idea generation, requirements collection, product planning, development, process planning, production, and operation, to disposal and recycling.

During the early phases of the product lifecycle (e.g. development phase) PDM systems are involved. PDM systems are used in the engineering and design communities in order to manage the product concept with the objective to generate reproducible product configurations. Therefore, PDM systems maintain the product design as versioned data in a repository. PDM functions as an interface between technical and commercial data processing, that is, between computer-aided design (CAD) systems on the one side, and procurement and production on the other side.

PLM systems extend PDM functionalities by managing the entire lifecycle of a product from conception to service and disposal. PLM integrates people, data, processes, and business systems, and provides a product information backbone for companies and their extended enterprise.

In the following we focus on two areas of engineering systems which are relevant for preservation:

- Data formats and models;
- Workflow, processes, and the product lifecycle.

### *Data Formats and Models*

**Data models.** A product is described from many different viewpoints, for example, -an electronic product is described by geometry, netlist, routing and placement, simulation models, 3D mechanical construction data, etc. Each different viewpoint is composed of a number of objects and sub-objects. Thus, we can formulate the following characteristics:

<sup>4</sup> International Organisation for Standardisation ISO9000: [http://www.iso.org/iso/iso\\_9000\\_essentials](http://www.iso.org/iso/iso_9000_essentials).

*Product data size.* The size of product data is large – the complete design information for a complex product may need gigabytes of data.

*Product data collection.* A product is described by a collection of digital objects. An electronic product is made up of many different components (enclosure, printed circuit boards, cables, software, etc.) which are developed in different systems that use different file formats. In addition, references to objects outside the boundaries of PLM systems may be involved, for example, to information captured in an enterprise resource planning (ERP) system (e.g. SAP R/3<sup>5</sup>).

*Multiple native formats.* Digital engineering objects are created in multiple native formats of CAD systems. Even one assembly in mechatronics might contain multiple produced design elements coming from different sources represented in several different file formats. A unified model covering all objects in all design and engineering disciplines does not exist.

*Data formats.* Design data is usually produced by the use of tools. Therefore, we can formulate two further characteristics which are important for digital preservation:

*Change of native formats.* Typically, CAD vendors release one major release per year which contains additional functionality. In most cases the internal data model changes, or is extended. Additionally, minor releases (service packs) are deployed several times a year, which combine bug fixes with minor enhancements.

*Copyright on formats.* Most of the native formats are proprietary and thus not generally readable. The interpretation of a native format usually requires the use of a specific tool of a specific vendor.

*References to other data.* Even in the context of a PLM system, the data may not be self-contained, that is there may exist external references, for example, to globally-maintained ontologies and classifications, or to objects in databases of collaboration partners. This leads to the two following characteristics:

*Proprietary taxonomy description.* Drawings, components, connectors, standard parts, and other regularly-used design elements are managed in specific libraries that are created and maintained external to a company. Global element libraries are mapped into local company libraries. This happens in many companies with proprietary taxonomy descriptions, including the geometrical layout of the design objects.

*Links to externally-developed designs.* Due to the complex nature of product designs, products are often created in collaborative projects by several companies. These companies are geographically separated, and each maintain their local design repositories that contain references to other parts of the product data model.

### ***Workflow, Processes, and the Product Lifecycle***

*Organisation of processes.* A product is developed in complex processes which partly follow strict workflows but which also contain creative phases of collaboration. These processes are organised in many different ways:

<sup>5</sup> SAP AG software corporation: <http://www.sap.com/about/company/history/1982-1991/index.epx>.



*Company-specific design methodologies.* The design methodology can vary significantly from company to company. In addition, the distinction must be made between companies which have a full-blown design process with electronic (chip, printed circuit board (PCB) , and cabling) and mechanical design, and other companies who have only a fraction of these.

*Project heterogeneity.* Even engineering projects within the same company can differ significantly regarding the policies and best practices they have to follow, depending on the kind of product, for which market and region, and so forth.

***Data and collaboration.*** Design processes are collaborative processes where a number of people work together to create products:

*Collaboration.* Product data is shared among people and tools. Collaboration takes place within the same company or across different companies. Inter-company collaboration requires sharing of information, but only as far as necessary for the collaboration. Companies are very sensitive to the protection of their data (intellectual property) from unwanted exchange with collaboration partners.

*Geographically-distributed product lifecycle.* Products are designed, manufactured, and serviced in a globally-distributed environment.

Many designs are variations of existing designs. Thus, existing designs have to be kept in a retrievable form:

*Design reuse.* Engineers often search, browse, and retrieve already-developed designs, in order to get ideas, examples, or reusable items.

The different design phases are performed in different contexts (tools, locations, people, etc.). Thus, a lot of metadata and context data are produced during the course of product development.

*Metadata and context data creation during the whole lifecycle.* During the design and development processes, not only product data but also metadata describing the performed processes are created. For example, during collaborations between different domains (e.g. electronic CAD, and mechanical CAD [eCAD/mCAD]), and during interactions between various PLM phases (e.g. a change request from requirement analysis to product development), social context data are created. This data needs to be collected during the whole product lifecycle.

*Version and configuration management system usage.* Existing PLM implementations contain version and configuration management that partly archive product data.

## Shortcomings and Problem Statement

Although several academic and industrial projects already tackle the problem area of long-term preservation of engineering data, some important aspects have not been considered so far (Brunsmann & Wilkes, 2009). The discussed characteristics represent a basis for finding missing project scopes and shortcomings of existing PLM implementations. The identified shortcomings are grouped by OAIS architecture functionality.

### ***General***

The state-of-the-art preservation feature in PLM-based systems is the creation and storage of versions in vast PLM repositories when certain milestones are reached. Archiving of product designs is done by creating backups of PLM repositories. Of course, these PLM repositories do not offer preservation functionalities. Thus, the change to a different tool or PLM system is not possible.

If data archiving is considered by existing projects at all, then only geometric data are archived. However, as described above, product designs do not consist only of geometric data, which means that the other data are lost forever.

### ***Data Ingest***

The practice of archiving and creating metadata only once makes the archival step quite complex. This lowers the acceptance of the archiving process. Unfortunately, methods for automatic extraction of metadata are currently missing during ingestion or during creation of data. Therefore, capturing of metadata is perceived as intrusive, and its value for later access to the archived data is not understood. In addition, valuable data and metadata are created during the whole lifecycle. Metadata include data about processes, people, products, provenance, and collaboration. If these data are not fully collected during the process, they are lost forever.

Company collaborations lead to distributed data repositories. The collection of distributed data for archiving is not provided by current PLM solutions. In addition, it is not guaranteed that all companies use tools of the same vendor in the same version.

The data that are created and ingested in one PLM phase are interconnected with objects from other phases, for example, a product design depends on the requirements document, a physical layout depends on the logical design, or a document created in the service phase depends on the product design. These relationships, and connections to other data and metadata that span over several PLM phases, have not been considered in long-term preservation projects in the engineering industry.

### ***Data Access***


For product design reuse and all other use cases, searching and finding of product designs is essential, but can only be done in a reasonable way if semantics are attached to the archived product data. Unfortunately, semantics are neither attached nor archived for a product design by existing PLM implementations.

Since the archived product designs are only available in native formats, the data cannot be used with future tool revisions unless the data are transformed into vendor-neutral formats.

### ***Preservation Planning***

Existing projects concentrate on migration as their preservation method, which creates problems due to the high frequency of format or system changes. Other existing projects use the method of transformation to vendor-neutral formats (Ball, Patel, & Ding, 2008; Lubell, Rachuri, Mani, & Subrahmanian, 2008). But transformation does not solve all problems, since the neutral format usually does not cover 100% of the native format, and vendor-neutral formats evolve with the native





formats. Other methods, like emulation or system preservation, have not been considered in more detail.

If the system environment and the proprietary product data are archived, it is not guaranteed that valid software licenses for applications are available if the archived native data are accessed several years later.

Since global taxonomy standards evolve, independently archived data that reference these global standards may become unusable without notice.

## Requirements

The examination of characteristics and shortcomings, and the analysis of the use case scenarios of the previous sections lead to a number of high-level requirements for digital preservation systems in PLM:

### *General*

*Process integration.* Since in all PLM phases relevant data are created, it should be possible to integrate archiving system functionality into all PLM-based workflows.

*Modularity and adaptability.* Since the design methodology is very different in different companies, a digital preservation system should be modular, customisable, and adaptable.

*Service interface availability.* It must be possible to access the preservation functionality via application programming interfaces. This ensures that tools written in different programming languages, and running on different operating systems, can access preservation functionality.

*System autonomy.* The preservation system has to be independent from the specific PLM system that currently uses the preservation system. It is desirable to enable a switch to another PLM system without loss of archived product data. It should also be possible to preserve the data without having access to an existing PLM system.

### *Data Ingest*

*Data and metadata, and file format standards.* The system should be able to transform multiple native formats into standard vendor-neutral formats. The system should be aware of standardised data and metadata and file formats so that future tools are able to interpret the data. The reference model should be able to accommodate existing standard metadata and file formats that are used by many companies and during collaborations. In addition, the model should allow new standards to be included.

*Parallel archiving of native and standard format.* The system must be able to archive both native and standard formats in parallel.

*Metadata generation and preservation.* To enable engineers to find existing designs which can be reused, the system should create and store metadata that describe the archived designs for searching on a reasonably detailed level. Metadata include graphics and significant properties. The extraction of metadata should be possible both automatically and manually.

*Process history tracking.* For audit and error-checking processes, it is necessary to archive not only the product data, but also information about the design process of a product. Therefore, at all stages of the design and preservation process, automated, as well as interactive, user actions should be monitored and added to the change history of the stored design objects. This practice enables users to search, retrieve, and understand a product design.

*Data validation.* The system should validate the data before they are ingested into the archive, and after the access from the archive. Validation should include checks for completeness and for correctness.

*Packaging of design data and metadata.* In order to allow the retrieval of parts of a product design (e.g. the motherboard of a computer), or data reflecting a certain stage of the design cycle (e.g. logical design or physical design), the system should be able to package both types of design data according to their type.

### **Data Access**

*Intellectual property rights protection.* The system should restrict the access to all CAD objects and properties in the design files. This ensures the protection of intellectual property that is required in collaboration processes where different companies are involved.

*Data validation.* The system should also validate the accessed data for completeness and correctness.

*Access to parts of a product design.* A product design consists of a collection of objects. The system should allow access to the complete design, as well as to parts of the design (e.g. logic design or mechanical construction data). If product data are reused with old design tools, it must be ensured that the design or directory structure is maintained in the archive.

*Search and retrieval of product data by metadata.* Engineers will require access to data from different projects in different data formats, created by different design tools. The system should provide discovery mechanisms for archived product designs.

### **Data Preservation Planning**

*Migration support.* Due to the frequent changes of file formats, the system should support the migration of archived product design in order to be in sync with the appropriate CAD tool and PLM system versions. Archived data must be migrated, either on access, or following a scheduled preservation plan.

*PLM process triggers for preservation planning activities.* Process milestones influence the necessity to keep specific formats in the archive, for example, after milestone “end of life”, the native format may become obsolete.

*External taxonomy dependencies.* Elements of component libraries are described by taxonomies that are external to the companies. The system should maintain references and mappings to specific versions of these taxonomies.

*Versioning of parts of a complete product design.* A product design consists of a hierarchy of dependent objects. However, it is possible that parts of an archived

product design are deleted or updated, or new objects are added to it (e.g. during the maintenance phase of a product). Therefore, the system should allow the creation of versions of a product design and put them under configuration control.

*External object maintenance.* The system should maintain relationships to objects outside of the PLM system, for example, to an ERP system.

## High-level System Architecture

In this section, we map the given requirements to the OAIS reference architecture, and derive extensions or adaptations to OAIS functionality. We identify steps in the workflows managed by PLM systems where a link to such a preservation system is possible.

As already described in the introduction, the SHAMAN project investigates the needs of three different application domains. These different domains share a similar lifecycle, which is shown in Figure 1, and described below.

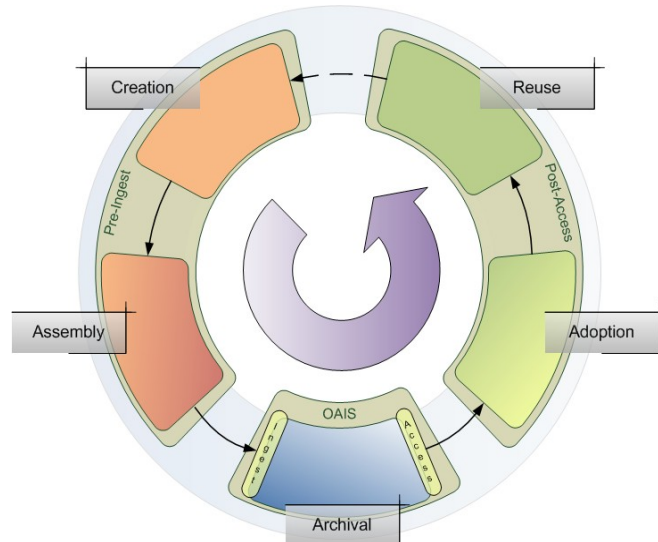


Figure 1. SHAMAN Information Lifecycle.

### *SHAMAN Information Lifecycle*

The proposed SHAMAN lifecycle (Brocks, Kranstedt, Jäschke, & Hemmje, [2009](#)) contains the phases pre-ingest with creation and assembly, archival, and post-access with adoption and reuse.

- During *pre-ingest*, all activities are executed that must be taken prior to the ingestion of data into the preservation system.
- The *creation* phase gives birth to new information that could be the result of complex processes involving many producers. In the engineering industry, this is the creation of archives of relevant metadata in real-time, during all phases of the PLM process.
- During the *assembly* phase, additional information that is relevant for archiving is collected. *Assembly* assures that enough information is archived, so that a digital object remains reusable for a future consumer. In the engineering industry, all relevant data are collected that are needed to

- correctly interpret a product design (i.e. geometry, coordinate list, metadata, product component libraries, simulation, and verification data).
- During the *archival* phase, the digital object is stored and maintained in a preservation system. Policies describe the lifetime and migration of digital objects.
  - The *post-access* phase comprises all activities that are needed for preparing the final access of the preserved data.
  - During the *adoption* phase, the archived digital objects are adapted for domain-specific reuse that a preservation system cannot accommodate. In the engineering industry, the archived data formats are translated to formats that are interpretable with current tools.
  - During the *reuse* phase, the consumer exploits the archived information. Reuse may lead to the creation of new digital objects that later are also candidates for archiving. In the engineering industry, an archived product design can be imported into a proprietary tool for creating product variations.

After having described the information lifecycle in general, we are able to specify a SHAMAN-based high-level system architecture for the engineering industry (see Figure 2). The systems are interconnected with solid (use-relationship) and dashed (reference-relationship) arrows. The relevant key aspects of the figure are described starting from the top.

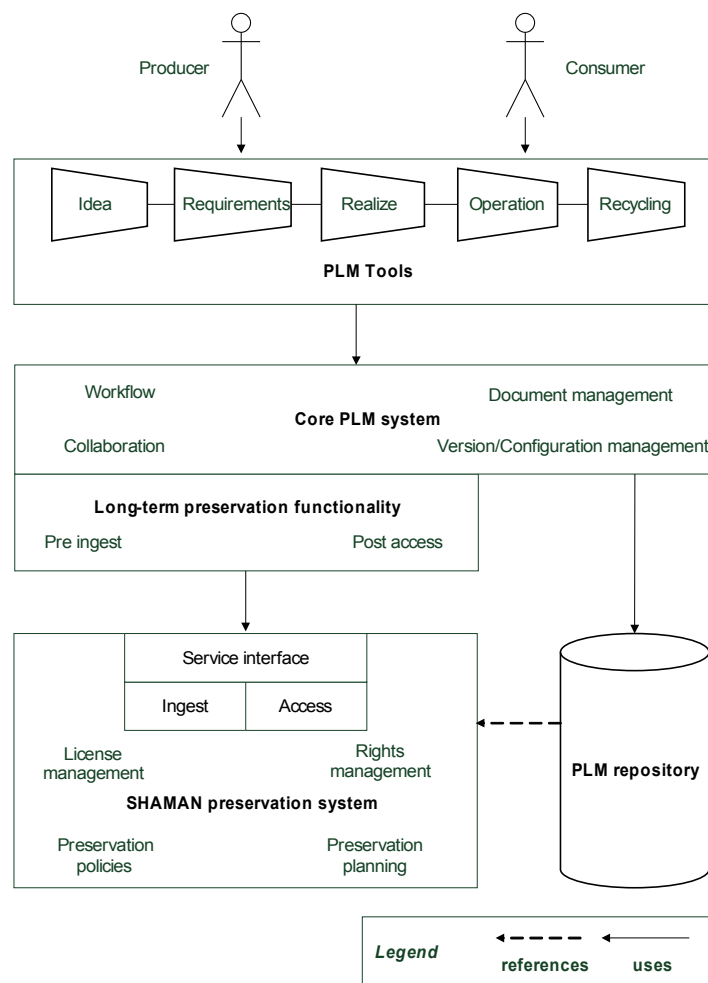


Figure 2. System Architecture.



### ***Actors and Tools***

During the whole product lifecycle (from idea generation to recycling) the system is used by producers and consumers, who are assisted by various tools that can be plugged into the overall PLM system infrastructure. The tools are able to display and modify the product data.

### ***Core PLM System***

The core PLM system consists of the standard PLM components, such as workflow, collaboration, and the management of data, versions, and configurations. The core PLM system will be used by the upper tool layer, and makes use of the long-term preservation functionality and the PLM repository that are both described below.

### ***Preservation-enabled PLM System***

The preservation-enabled PLM system consists of the core PLM system enhanced with preservation functionality. This preservation functionality is responsible for executing relevant actions during the pre-ingest and post-access phase.

For example, during the pre-ingest phase the system collects all relevant product data that are needed for archiving. Also, during pre-ingest, native, proprietary formats are converted into vendor-neutral formats for archiving. In addition, it may also be decided to archive the system environment (software, tools, etc.). Pre-ingest rules determine what and when to archive. For example, if a specific phase of the PLM process is reached, a rule might determine to archive the product design.

During post-access, the preservation-enabled PLM system transforms the vendor-neutral archival format into the current native format. Also, the system might reconstruct the previously archived system environment. In addition, domain-specific *post-access rules* are specified and executed. A post-access rule might embed the retrieved product design into the current tool landscape, and create relevant relationships between the product design and current data models and ontologies.

Pre-ingest and post-access rules are domain-specific, and do not include activities that are part of the preservation.

### ***SHAMAN Preservation System***

The preservation-enabled PLM system makes use of the SHAMAN preservation system. This preservation system offers a service interface that includes ingest and access functionality. Such a service interface decouples the preservation system from other systems, and allows flexible use, even from systems that reside in other companies or institutions. Therefore, it is necessary that the preservation system also includes the management of rights, to prevent unauthorised access to the archived data. The preservation system also includes standard preservation planning functionality (obsolescence detection) and preservation policy management. The management of software licenses is also needed to ensure legal access to file formats and tools.

*Preservation policies* define principles that guide the preservation of digital objects according to company, project, document type, or even object-based guidelines. The PLM system uses such policies during ingest in order to define the desired handling of ingested product data. Policies are needed because engineering projects differ significantly regarding the kind of product, market, and country. For

example, if the collection of ingested product data consist of geometric, logical and physical designs, a preservation policy is able to specify that legal demands allow the deletion of archived physical designs after an elapsed amount of time. Preservation policies should guarantee that a preservation system is self-sustaining and that the archived content remains in a usable state.

### ***PLM Repository***

The PLM repository stores the native product data and is the operational database for the core PLM system. It is possible that the PLM repository references the content of the preservation system. For example, if parts of a design move into the preservation system, then the PLM repository can store a unique reference to the archived portion of the design, and remove the archived product design from the PLM repository. If this design is accessed later, the PLM system will retrieve the design from the preservation system. If the archived data are migrated then references must be kept in a valid state. References from the preservation system to the PLM repository may not exist, since the preservation system must be independent from systems that use the preservation functionality.

### ***Summary***

Since the architecture was derived from requirements, it is worthwhile to verify if the key aspects of the architecture are a good match to the given requirements.

*General.* The preservation system is modular, and has a service interface that allows the integration of the system in all phases of the PLM process.

*Data ingest.* During pre-ingest, relevant metadata (e.g. process history) are collected, and might be transformed into vendor-neutral formats before ingest. Also, during pre-ingest, the data can be validated for completeness. Due to the parallel usage of a preservation system and a PLM repository, both native and transformed data are kept.

*Data access.* During post-access the data are validated for completeness. The service interface of the preservation system allows users to search via metadata. Intellectual property rights are protected by the preservation system, and it controls the access to parts of the product design.

*Data preservation planning.* The preservation system migrates the archived data whenever they are needed, and during pre-ingest, the preservation-enabled PLM systems collect all relevant product design data.

## **Conclusion and Outlook**

Based on PLM system characteristics we have described extensions that are needed for integration of preservation functionality into PLM workflows. We also have looked at requirements for a preservation system based on the needs of the Designated Community in the engineering industry. One of the major requirements is easy integration of digital preservation processes and PLM workflows by using configurable modular solutions. This requirement is met by a PLM system architecture that collaborates with the functionality of a SHAMAN-based preservation system while conserving the possibility of executing the regular PLM processes.



Further investigations are needed to specify the service interface in more detail. The consequences of distributed archives resulting from global cross-company collaborations in the engineering industry must be reflected upon. It remains to be investigated which relevant metadata in PLM processes exist, and how these metadata can be captured, archived and preserved. Future research will also tackle the problem of maintaining dependencies to external taxonomies that are used in the engineering industry. The problem of maintaining version-management metadata has to be addressed, since such version metadata are currently held in the PLM system. If version information should be accessible by other PLM systems, version metadata has also to be archived and maintained in a preservation system.

## Acknowledgements

This paper is supported by the European Union within the 7<sup>th</sup> Framework Programme Integrated Project: SHAMAN.

## References

- Ball, A., Patel, M., & Ding, L. (2008). Towards a curation and preservation architecture for CAD engineering models. In *Proceedings of the 5<sup>th</sup> International Conference on Preservation of Digital Objects*. London, UK. Retrieved 20 February, 2011, from [http://www.bl.uk/ipres2008/presentations\\_day1/17\\_Ball.pdf](http://www.bl.uk/ipres2008/presentations_day1/17_Ball.pdf).
- Brocks, H., Kranstedt, A., Jäschke, G., & Hemmje, M. (2010). *Modeling Context for Digital Preservation*. In Szczerbicki, E., & Nguyen, N. (Eds.) *Studies in Computational Intelligence: Vol. 260. Smart Information and Knowledge Management* (p.197-226). Berlin: Springer.
- Brunsmann, J., & Wilkes, W. (2009). *State-of-the-art of long-term archiving in product lifecycle management*. Manuscript submitted for publication.
- Consultative Committee for Space Data Systems (2002). *Reference model for an open archival information system*. ISO 14721. Retrieved February 20, 2011, from <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- Heutelbeck, D., Brunsmann, J., Wilkes W., & Hundsörfer, A. (2009). Motivations and challenges for digital preservation in design and engineering. In *Proceedings of 1<sup>st</sup> International Workshop on Innovation in Digital Preservation, June 19, 2009*. Austin, TX. Retrieved February 20, 2011, from <http://cs.harding.edu/indp/papers/heutelbeck5.pdf>.
- Lubell, J., Rachuri, S., Mani, M., & Subrahmanian, E. (2008). Sustaining engineering informatics: Towards methods and metrics for digital curation. *The International Journal of Digital Curation* 3, (2).
- SHAMAN (2008). *Survey of users and providers in Europe and SHAMAN usage scenarios*. SHAMAN Internal Project Deliverable 1.1. <http://www.shaman-ip.eu:8000/SHAMAN/wiki/WP1> (access restricted to SHAMAN Consortium Members).